A Scalable Design of Experiments Framework for Optimal Sensor Placement

Jing Yu[¶], Victor M. Zavala[†], and Mihai Anitescu[¶,‡]

[¶]Department of Statistics
University of Chicago, 15734 S. University Avenue, Chicago, IL 60637, USA

[†]Department of Chemical and Biological Engineering
University of Wisconsin-Madison, 1415 Engineering Dr., Madison, WI 53706, USA

[‡]Mathematics and Computer Science Division
Argonne National Laboratory, 9700 S. Cass Ave., Lemont, IL 60439, USA

Abstract

We present a scalable design of an experiments framework for sensor placement in the context of state estimation of natural gas networks. We aim to compute optimal sensor locations where observational data are collected, by minimizing the uncertainty in parameters estimated from Bayesian inverse problems, which are governed by partial differential equations. The resulting problem is a mixed-integer nonlinear program. We approach it with two recent heuristics that have the potential to be scalable for such problems: a sparsity-inducing approach and a sum-up rounding approach. We investigate two metrics to guide the design of experiments (the total flow variance and the A-optimal design criterion) and analyze the effect of different noise structures (white and colored). We conclude that the sum-up rounding approach produces shrinking gaps with increased meshes. We also observe that convergence for the white noise measurement error case is slower than for the colored noise case. For A-optimal design, the solution is close to the uniform distribution, but for the total flow variance the pattern is noticeably different.

1 Introduction

The sensor placement problem seeks to determine the optimal number, locations, and types of sensors that would maximize information about a dynamical system, where information is often expressed in terms of the posterior covariance matrix of the states of parameters of the system. The problem can thus be cast as an optimal design of the experiments problem. Such a problem is computationally challenging, particularly in the infinite-dimensional case, because one must solve a mixed-integer and bilevel optimization problem constrained by differential algebraic equations or by partial differential equations (PDEs). This problem has been addressed by using mixed-integer programming techniques, for contaminant detection in water networks [3, 2, 22, 10]. In these studies, an optimal set of sensor locations is selected from a set of candidate locations to minimize a certain engineering

metric such as contaminant detection time, population exposure, or likelihood of detection. Likelihoods are assigned based on contamination scenarios, and not on information content of the sensor data recorded, as in a traditional experimental design setting. These problems, however, fail to provide statistically meaningful sensor network designs. Moreover, because the formulations capture flow dynamics by using surrogate representations such as transportation delays, they fail to capture nonlinear effects.

Sensor placement problems also have been addressed in a more general control setting where one seeks to optimize a measure of observability such as the covariance matrix, Kalman estimator gain, or so-called observability Grammian matrix. This problem is again a bilevel optimization problem. The covariance matrix approach in [6] bypasses this by assuming that the dynamic model is linear, thus allowing the inner minimization problem to be formulated as a linear matrix inequality. The approach in [20] models the dynamics of the covariance matrix directly as a Riccati differential equation, which implicitly assumes linearity and thus enables the use of semidefinite programming algorithms. However, the work in [20] is focused on control policy design to extract maximum information, and not on sensor placement design. Consequently, the authors do not consider mixed-integer formulations. A rigorous treatment of nonlinear dynamics is presented in [19] by casting the problem as a mixed-integer nonlinear program. The authors use a genetic algorithm to deal with the inner minimization problem that computes the observability metric. A similar approach is used in [14] to address the inner minimization problem. Mixed-integer techniques have also been used in the context of information maximization for Gaussian processes and for designing Latin hypercube samples [11, 8]. These approaches do not capture physical models.

Recently, the sensor placement problem for systems described by PDEs has been cast as an A-experimental design problem in which the number of sensors (i.e., the design cost) is controlled by using an ℓ_0 regularization norm that is in turn approximated by using a smoothing function [1]. This approach was shown to be scalable and applicable to infinite-dimensional systems, but it requires tuning and can be numerically unstable. One can also formulate and solve the problem as a mixed-integer programming problem directly, but this approach can become computationally intractable because the PDEs are in general nonconvex and because the problem is bilevel.

In this work we present a design of an experiments framework for sensor placement that seeks to compute optimal sensor locations where observational data are collected, by minimizing the uncertainty in parameters estimated from Bayesian inverse problems, which are governed by PDEs. The resulting problem is a mixed- integer infinite-dimensional optimal control problem. We approach this problem using two efficient heuristics that have the potential to be scalable for such problems: a sparsity-inducing approach [1] and a sum-up rounding approach [17]. We investigate two objectives: the total flow variance and the A-optimal design criterion. We conclude that the sum-up rounding approach produces shrinking gaps with increased meshes. We also observe that convergence for the white noise measurement error is slower than for the colored noise case. For the A-optimal design the solution is close to the uniform distribution, but for the total flow variance the pattern is noticeably different.

An important application of optimal sensor location techniques is infrastructure networks (oil, water, gas, and electricity) in which large amounts of sensor data need to be processed in real time

in order to reconstruct the state of the system or to identify leaks, faults, or attacks. In this work we focus on natural gas networks, which are used to transport fuel to power generation facilities and urban areas from storage and processing facilities. These networks comprise pipelines that span thousands of miles and exhibit complex dynamics. An interesting property of natural gas networks is that significant amounts of gas can be stored inside the pipelines. The stored gas is distributed spatially along the pipelines and is normally referred to as *line-pack* [5]. Line-pack is used by pipeline operators to modulate variations of gas demands at multiple spatial points in intraday operations. Some of the strongest variations in gas demands are the result of on-demand startup and shutdown of gas-fired power plants [15]. Modulating these variations is challenging because the fast release of line-pack at multiple simultaneous locations can trigger complex spatiotemporal dynamic responses that propagate hundreds to thousands of miles and that can take hours to stabilize. Therefore, line-pack management is performed by using sophisticated optimal control and pipeline simulation tools. To use these automation tools, one must reconstruct spatiotemporal state fields (flows, pressures, temperatures) [16] and natural gas leaks [4]. This task is challenging from a practical stand point given the limited amounts of sensor data (often limited to pressure and flow signals at a finite set of locations), the infinite-dimensional nature of pipeline systems, and the complex physical behavior of these systems. Such challenges are not unique to natural gas networks but also arise in other domains such as geophysics and contaminant source detection in water networks.

The paper is structured as follows. In §2 we define the physical system model that we use in designing the experiments. In §3 we provide the formulation of the experiments problems we aim to solve. In §4 we present numerical experiments with machine learning and the sum-up rounding procedure for solving the design of experiments problem. In §5 we summarize our conclusions and briefly describe future work.

2 Distributed System Modeling

We illustrate the complexity of the optimal sensor placement problem by focusing on the physical equations describing the dynamics of natural gas networks. Details on the model derivation, nomenclature, and units used in this section can be found in [24].

2.1 Problem Physics

The isothermal flow of gas through a horizontal pipeline is described by the conservation and momentum equations:

$$\frac{\partial \rho(\tau,x)}{\partial \tau} + \frac{\partial (\rho(\tau,x)\nu(\tau,x))}{\partial x} = 0 \tag{2.1a}$$

$$\frac{\partial (\rho(\tau,x)\nu(\tau,x))}{\partial \tau} + \frac{\partial p(\tau,x)}{\partial x} = -\frac{\lambda}{2D}\rho(\tau,x)\nu(\tau,x)|\nu(\tau,x)|. \tag{2.1b}$$

Here, $\tau \in \mathcal{T} := [0,T]$ is the time dimension with final time T (planning horizon), and $x \in \mathcal{X} := [0,L]$ is the axial dimension with length L. The pipeline diameters are denoted as D, and the friction coefficients are denoted as λ. The states of the link are the gas density $\rho(\tau,x)$, the gas speed $\nu(\tau,x)$,

and the gas pressure $p(\tau, x)$. The transversal area A, volumetric flow $q(\tau, x)$, and mass flow $f(\tau, x)$ are given by

$$A = \frac{1}{4}\pi D^2 \tag{2.2a}$$

$$q(\tau, x) = \nu(\tau, x)\, A \tag{2.2b}$$

$$f(\tau, x) = \rho(\tau, x)\, \nu(\tau, x)\, A. \tag{2.2c}$$

For an ideal gas, pressure and density are related as follows:

$$\frac{p(\tau, x)}{\rho(\tau, x)} = c^2. \tag{2.3}$$

Here, c is the gas speed of sound. The speed (assuming an ideal gas behavior) and the friction factor λ can be computed from

$$c^2 = \frac{\bar{\gamma} Z R T}{M} \tag{2.4a}$$

$$\lambda = \left(2 \log_{10}\left(\frac{3.7\, D}{\epsilon}\right)\right)^{-2}, \tag{2.4b}$$

where Z is the gas compressibility factor, R is the universal gas constant, T is the gas temperature, M is the gas molar mass, ϵ is the pipe rugosity, and $\bar{\gamma}$ is the adiabatic constant.

Often one desires to transform (2.1) into a more convenient form in terms of mass flow and pressure by using (2.3) and (2.2):

$$\frac{\partial p(\tau, x)}{\partial \tau} + \frac{c^2}{A}\frac{\partial f(\tau, x)}{\partial x} = 0 \tag{2.5a}$$

$$\frac{1}{A}\frac{\partial f(\tau, x)}{\partial \tau} + \frac{\partial p(\tau, x)}{\partial x} = -\frac{\lambda \rho(\tau, x)}{2D}\frac{f(\tau, x)}{\rho(\tau, x)\, A}\left|\frac{f(\tau, x)}{\rho(\tau, x)\, A}\right|. \tag{2.5b}$$

Substituting (2.3) and (2.2a) in (2.5b) and performing some manipulations, we obtain the more compact form:

$$\frac{\partial p(\tau, x)}{\partial \tau} = -\frac{c^2}{A}\frac{\partial f(\tau, x)}{\partial x} \tag{2.6a}$$

$$\frac{1}{A}\frac{\partial f(\tau, x)}{\partial \tau} = -\frac{\partial p(\tau, x)}{\partial x} - \frac{8\,\lambda\, c^2}{\pi^2\, D^5}\frac{f(\tau, x)|f(\tau, x)|}{p(\tau, x)}. \tag{2.6b}$$

For numerical purposes, we define scaled flows $f(\tau, x) \leftarrow \alpha_f f(\tau, x)$ and pressures $p(\tau, x) \leftarrow \alpha_p p(\tau, x)$, where α_f and α_p are scaling factors. Scaling (2.6) and rearranging, we obtain the final form:

$$\frac{\partial p(\tau, x)}{\partial \tau} = -c_1\frac{\partial f(\tau, x)}{\partial x},\ \tau \in \mathcal{T}, x \in \mathcal{X} \tag{2.7a}$$

$$\frac{\partial f(\tau, x)}{\partial \tau} = -c_2\frac{\partial p(\tau, x)}{\partial x} - c_3\frac{f(\tau, x)|f(\tau, x)|}{p(\tau, x)},\ \tau \in \mathcal{T}, x \in \mathcal{X}, \tag{2.7b}$$

where the constants c_1, c_2, and c_3 are given by

$$c_1 = \frac{\nu^2}{A}\frac{\alpha_p}{\alpha_f}, \quad c_2 = A\frac{\alpha_f}{\alpha_p}, \quad c_3 = \frac{8\,\lambda\,\nu^2\,A}{\pi^2\,D^5}\frac{\alpha_p}{\alpha_f}. \tag{2.8}$$

For subsonic flow, one must impose a boundary condition at the inlet point and a boundary condition at the outlet point. For instance one can specify pressure at the inlet and outlet points,

$$p(0,\tau) = \theta^{orig}(\tau) \tag{2.9a}$$
$$p(L,\tau) = \theta^{rec}(\tau). \tag{2.9b}$$

One also can impose boundary conditions for inlet and outlet flows as

$$f(0,\tau) = f^{orig}(\tau) \tag{2.10a}$$
$$f(L,\tau) = f^{rec}(\tau). \tag{2.10b}$$

Alternatively, one can impose a boundary condition for pressure at the inlet point and one for flow at the outlet point, or vice versa.

2.2 Discretization of State Equations

For either simulation or optimization we need to discretize equations (2.7). These equations are a particular case of a nonlinear system of equations. To that end, we introduce the vector variable

$$u(t,x) = \begin{pmatrix} p(t,x) \\ f(t,x) \end{pmatrix}, \quad (t,x) \in [0,T] \times [0,L]$$

which consists of pressure $p(t,x)$ and flow $f(t,x)$ in the system defined over the domain: the Cartesian product of $[0,T]$ in time with $[0,L]$ in space. With this notation, the governing equations of a gas pipeline can be written as the following nonlinear system of PDEs:

$$\frac{\partial u}{\partial t} + \begin{bmatrix} 0 & c_1 \\ c_2 & 0 \end{bmatrix} \frac{\partial u}{\partial x} + c_3 \begin{bmatrix} 0 \\ f|f|/p \end{bmatrix} = \mathbf{0}. \tag{2.11}$$

The parameters $c_1, c_2, c_3 > 0$ are defined in (2.8) and play a key role in identifying stable numerical schemes for solving (2.11). The initial conditions are given by $p(0,x) = p_0(x)$ and $f(0,x) = f_0(x)$. We use prescribed and constant pressure boundary conditions: $p(t,0) = p_1$ and $p(t,L) = p_2$. We note that experimental validation has indicated that constant pressure boundary conditions are appropriate for gas pipeline systems [21].

We now discretize the system of PDEs (2.11). The system (2.11) is not conservative, since the friction term (nonlinear term) results in dissipation of energy. The linear part of the system (formally obtained by setting c_3 to 0) represents a conservative hyperbolic system, since all the eigenvalues of the 2×2 matrix in (2.11) are real and equal to $\pm\sqrt{c_1 c_2}$. At each point, this system has two characteristic directions each having an angle smaller than 90-degrees with one of the boundaries. To maintain

stability, we use an upwinding scheme along each of the characteristics [12].

We first consider the linear part of the system:

$$\frac{\partial u}{\partial t} + B\frac{\partial u}{\partial x} = 0, \quad x \in [0, L], t \in [0, T]$$

with $B = \begin{pmatrix} 0 & c_1 \\ c_2 & 0 \end{pmatrix} \in \mathbb{R}^{2\times 2}$. B has eigenvalue decomposition $B = S\Lambda S^{-1}$, where

$$\Lambda := \text{diag}\{\lambda_1, \lambda_2\} = \begin{pmatrix} -\sqrt{c_1 c_2} & 0 \\ 0 & \sqrt{c_1 c_2} \end{pmatrix}, \quad S := \begin{pmatrix} \sqrt{c_1} & \sqrt{c_1} \\ -\sqrt{c_2} & \sqrt{c_2} \end{pmatrix}.$$

We define the characteristic variable $\tilde{u} := S^{-1}u$, which satisfies the decoupled system

$$\frac{\partial \tilde{u}}{\partial t} + \Lambda\frac{\partial \tilde{u}}{\partial x} = 0.$$

The system consists of two independent wave equations traveling in opposite directions. To separate the two waves, we introduce the splitting of the eigenvalues λ_k as

$$\lambda_k = \lambda_k^+ + \lambda_k^-, \quad \lambda_k^+ := \max(\lambda_k, 0), \quad \lambda_k^- := \min(\lambda_k, 0).$$

We can write the upwind scheme for the characteristic variable as

$$\frac{1}{\Delta t}(\tilde{u}_j^{n+1} - \tilde{u}_j^n) + \frac{1}{\Delta x}\Lambda^+(\tilde{u}_j^n - \tilde{u}_{j-1}^n) + \frac{1}{\Delta x}\Lambda^-(\tilde{u}_{j+1}^n - \tilde{u}_j^n) = 0$$

with

$$\Lambda^+ := \text{diag}(\lambda_1^+, \lambda_2^+), \quad \Lambda^- := \text{diag}(\lambda_1^-, \lambda_2^-).$$

Next we define $B^+ := S\Lambda^+ S^{-1}, B^- := S\Lambda^- S^{-1}$, and obtain the upwinding scheme in terms of the original variable $u(\cdot)$ by multiplying the resulting scheme by the matrix S,

$$\frac{1}{\Delta t}(u_j^{n+1} - u_j^n) + \frac{1}{\Delta x}B^+(u_j^n - u_{j-1}^n) + \frac{1}{\Delta x}B^-(u_{j+1}^n - u_j^n) = 0, \tag{2.12}$$

where

$$u_j^n = \begin{pmatrix} p_j^n \\ f_j^n \end{pmatrix} \quad B^+ = \frac{1}{2}\begin{pmatrix} \sqrt{c_1 c_2} & c_1 \\ c_2 & \sqrt{c_1 c_2} \end{pmatrix} \quad B^- = \frac{1}{2}\begin{pmatrix} -\sqrt{c_1 c_2} & c_1 \\ c_2 & -\sqrt{c_1 c_2} \end{pmatrix}.$$

Here, the notation u_j^n indicates the j-the point in the spatial mesh and the n-th point in the temporal mesh. Plugging these terms back into (2.12), we obtain the discretization scheme:

$$p_j^{n+1} = p_j^n + \frac{\Delta t}{2\Delta x}[\sqrt{c_1 c_2}(p_{j-1}^n - 2p_j^n + p_{j+1}^n)] - \frac{\Delta t}{2\Delta x}[c_1(f_{j+1}^n - f_{j-1}^n)] \tag{2.13}$$

$$f_j^{n+1} = f_j^n + \frac{\Delta t}{2\Delta x}[\sqrt{c_1 c_2}(f_{j-1}^n - 2f_j^n + f_{j+1}^n)] - \frac{\Delta t}{2\Delta x}[c_2(p_{j+1}^n - p_{j-1}^n)]. \tag{2.14}$$

Stability of this scheme is ensured if the corresponding scalar upwind schemes for all variables $\tilde{u}_k$

are stable, which gives the Courant, Friedrichs, and Lewy (CFL) stability condition:

$$\max_k \ |\lambda_k| \frac{\Delta t}{\Delta x} = \sqrt{c_1 c_2} \cdot \frac{\Delta t}{\Delta x} \leq 1.$$

Since we anticipate that the friction term will not dominate, we simply consider the upwinding scheme for each characteristic for the linear equation to which we add the friction term explicitly. This procedure results in the following numerical scheme:

$$p_j^{n+1} = p_j^n + \frac{\Delta t}{2\Delta x}[\sqrt{c_1 c_2}(p_{j-1}^n - 2p_j^n + p_{j+1}^n)] - \frac{\Delta t}{2\Delta x}[c_1(f_{j+1}^n - f_{j-1}^n)] \tag{2.15a}$$

$$f_j^{n+1} = f_j^n + \frac{\Delta t}{2\Delta x}[\sqrt{c_1 c_2}(f_{j-1}^n - 2f_j^n + f_{j+1}^n)] - \frac{\Delta t}{2\Delta x}[c_2(p_{j+1}^n - p_{j-1}^n)] - \Delta t \cdot c_3 f_j^n |f_j^n| / p_j^n. \tag{2.15b}$$

for $j = 1, 2, ..., N_x - 1$ and $n = 0, 1, ..., N_t$. The friction term can be split among the characteristic equations based on the eigenvectors of the matrix of the linear system and the boundary. To simplify the implementation, we repeat the flux values of the last interior node: $f_0^n = f_1^n,\ f_{N_x}^n = f_{N_x-1}^n$.

3 Design of Experiments Setup

3.1 Bayesian Framework

We consider a setting with measurements perturbed by additive Gaussian noise:

$$\mathbf{d} = \boldsymbol{f}(\boldsymbol{m}) + \boldsymbol{\eta}, \quad \boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}_{\text{noise}}), \tag{3.16}$$

where $\boldsymbol{\Gamma}_{\text{noise}} \in \mathbb{R}^{q \times q}$ is the measurement noise covariance matrix and $\boldsymbol{f}$ is a nonlinear operator that maps a parameter vector $\boldsymbol{m} \in \mathbb{R}^n$ to the space-time observation vector $\mathbf{d} \in \mathbb{R}^q$. Formally,

$$\boldsymbol{f} : \boldsymbol{m} \overset{\mathcal{S}}{\longmapsto} \bar{\boldsymbol{u}} \overset{\mathcal{B}}{\longmapsto} \mathbf{d} \tag{3.17}$$

where $\boldsymbol{m}$ are the input parameters, $\mathcal{S}$ is the discretized PDE solution operator, $\bar{\boldsymbol{u}} \in \mathbb{R}^s$ is the discretized PDE solution vector, and $\mathcal{B}$ is the state-to-observation operator. In our case, the input parameters $\boldsymbol{m}$ (the inferred variables) consist of the initial pressure and flow at the grid points:

$$\boldsymbol{m} = \left\{ \{p_0(i\Delta x)\}_{i=1,2,...,N_x-1}, \{f_0(i\Delta x)\}_{i=0,1,...,N_x} \right\}. \tag{3.18}$$

The solution vector consists of the pressure and flow at all nodes and all times:

$$\bar{\boldsymbol{u}} = \left\{ \{p_j^n\}_{j=0,1,2,...,N_x,\ n=0,1,2,...,N_t}, \{f_j^n\}_{j=0,1,2,...,N_x,\ n=0,1,2,...,N_t} \right\}. \tag{3.19}$$

The map $\mathcal{S}$ is defined by the numerical scheme (2.15). The observations $\mathbf{d}$ are a subset of entries in the solution vector $\bar{\boldsymbol{u}}$, and the *space-time observation operator* $\mathcal{B}$ is the restriction operator from the components of $\bar{\boldsymbol{u}}$ to the entries in $\mathbf{d}$. For our experimental design, we assume that the sensors are fixed and interrogated at all times, in which case the observation vector d and observation operator

$\mathcal{B}$ are parameterized only by the spatial locations at which we observe the pressure and flow.

The measurement noise $\boldsymbol{\eta}$ is independent of $\boldsymbol{m}$ and thus $\mathbf{d}|\boldsymbol{m} \sim \mathcal{N}(\boldsymbol{f}(\boldsymbol{m}), \boldsymbol{\Gamma}_{\text{noise}})$. The likelihood is given by

$$\pi_{\text{like}}(\mathbf{d}|\boldsymbol{m}) \propto \exp\Big(-\frac{1}{2}\|\boldsymbol{f}(\boldsymbol{m}) - \mathbf{d})\|^2_{\boldsymbol{\Gamma}^{-1}_{\text{noise}}}\Big). \tag{3.20}$$

Stating the consequence of Bayes' theorem $\pi_{\text{post}}(\boldsymbol{m}|\mathbf{d}) \propto \pi_{\text{like}}(\mathbf{d}|\boldsymbol{m})\pi_{\text{prior}}(\boldsymbol{m})$ with a Gaussian prior $\pi_{\text{prior}}(\boldsymbol{m}) \propto \exp\Big(-\frac{1}{2}\|\boldsymbol{m} - \boldsymbol{m}_{\text{prior}}\|^2_{\boldsymbol{\Gamma}^{-1}_{\text{prior}}}\Big)$, we obtain the parameterization of the posterior distribution $\pi_{\text{post}}(\boldsymbol{m}|\mathbf{d})$ (up to a constant) as [9]:

$$\pi_{\text{post}}(\boldsymbol{m}|\mathbf{d}) \propto \exp\Big(-\frac{1}{2}\|\boldsymbol{f}(\boldsymbol{m}) - \mathbf{d})\|^2_{\boldsymbol{\Gamma}^{-1}_{\text{noise}}} - \frac{1}{2}\|\boldsymbol{m} - \boldsymbol{m}_{\text{prior}}\|^2_{\boldsymbol{\Gamma}^{-1}_{\text{prior}}}\Big), \tag{3.21}$$

where $\boldsymbol{m}_{\text{prior}}$ is the mean of the prior distribution, $\boldsymbol{\Gamma}_{\text{prior}} \in \mathbb{R}^{n\times n}$ is the covariance matrix for the prior which we assume to be a scalar multiple of the identity matrix. $\boldsymbol{\Gamma}_{\text{noise}} \in \mathbb{R}^{q\times q}$ is the covariance matrix for the noise. We consider two types of noise matrices $\boldsymbol{\Gamma}_{\text{noise}}$. The first one is $\text{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_q^2)$ which indicates independent measurements (white noise). The second one is given in (3.22), which assumes independence in space but nonzero and decaying correlation in time (colored noise). The statistical parameters needed to define our model are $\boldsymbol{\Gamma}_{\text{noise}}$, $\boldsymbol{\Gamma}_{\text{prior}}$, and $\boldsymbol{m}_{\text{prior}}$.

By measurement noise we understand here the discrepancy between simulated pressure and flow given exact initial and boundary information and the output of the sensor of a fixed location. For modeling this measurement noise we have several considerations. The intrinsic sensor errors can be assumed to be statistically independent between different sensors. However, some of the discrepancy between sensor indications and computed flow are also due to the numerics and boundary conditions, such as unresolved fluctuations and external perturbations. If the sensors are sufficiently far apart, we can assume that insofar as these are represented as probabilistic errors, they are spatially independent. The situation concerning the temporal features of the noise is more complicated, however. The discrepancy between measurement and simulation can be due to unresolved scales, which typically have nonzero correlation times and cannot be ignored if the measurements are frequent enough. We thus model the measurement noise as a Gaussian random variable that is independent in space but correlated in time. We also assume that its mean is zero. For the intrinsic error of a calibrated sensor, this is a reasonable assumption. Given our definition, measurement noise also includes numerical error. We assume that this and all other biases are small enough compared with the sensor error (and given the optimal variances, this is a reasonable assumption). In summary we assume that the measurement error has zero mean and covariance given by the following function:

$$\text{Cov}\big((t_i, x_i), (t_j, x_j)\big) = \delta(x_i, x_j) \exp\left\{-\frac{|t_i - t_j|}{\tau_i} I_{\{|t_i - t_j| \le \tau_j\}}\right\}, \tag{3.22}$$

where τ_i, τ_j are parameters with dimension of time that define the shape of the covariance function. Here $\delta(x_i, x_j)$ is the Kronecker δ symbol, which takes the value 1 if $x_i = x_j$ and 0 otherwise. Γ_{noise} is then computed by evaluating the covariance functions at the position where pressure and flow are measured. In other words, entries on subdiagonals of Γ_{noise} have exponential decay, and Γ_{noise} is a sparse matrix.

We also experiment with white noise in time. This is probably the most common usage in such problems even if it does not make sense in the limit of dense temporal observations. It does, however, has the advantage of needing fewer parameters. Moreover, under proper scaling conditions, this gives a conservative approximation of the variance for target linear forms of the initial state (i.e., it overestimates the posterior variance). We thus consider the case of white noise in time and space as well, which corresponds to a constant diagonal covariance function, that is:

$$\mathrm{Cov}\big((t_i, x_i), (t_j, x_j)\big) = \delta(x_i, x_j)\delta(t_i, t_j). \tag{3.23}$$

The other element in defining a Bayesian uncertainty framework concerns the prior assumptions about the parameters to be inverted, $\boldsymbol{m}$. Here we use a Gaussian prior, which is a common choice for Bayesian inverse problems [9]. The prior mean describes our best guess about the uncertainty parameter, which can be obtained from historical measurements or from other available information. In addition, because of the lack of a priori information about the parameters, we will use a prior that assumes spatial independence. This setup can be interpreted as making no assumptions about the smoothness of the initial data, which should result in conservative statements.

Despite the choice of Gaussian prior and noise probability distributions, the posterior probability distribution need not be Gaussian, because of the nonlinearity of $\boldsymbol{f}(\boldsymbol{m})$ [9]. If our purpose were estimation, then we would aim to characterize the posterior distribution. The mean of this posterior distribution, $\boldsymbol{m}_{\text{MAP}}$, is the parameter vector maximizing the posterior (3.21), and is known as the *maximum a posteriori* (MAP) point. It can be found by minimizing the negative log posterior, which amounts to solving the following optimization problem:

$$\boldsymbol{m}_{\text{MAP}} = \arg\min_{\boldsymbol{m}} \mathcal{J}(\boldsymbol{m}) := -\log \pi_{\text{post}}(\boldsymbol{m}|\mathbf{d}). \tag{3.24}$$

Characterizing the posterior uncertainty, however, would require exploring and summarizing this posterior distribution, which in general can be done only with Markov-chain Monte Carlo methods [9]. To simplify computations, we use Laplace's approximation. That is, we make a quadratic approximation of the negative log of the posterior (3.21) around the MAP point. The posterior covariance matrix $\boldsymbol{\Gamma}_{\text{post}}$ is then given by the inverse of the Hessian of $\mathcal{J}$ at $\boldsymbol{m}$. We thus approximate the posterior covariance with a Gaussian distribution, with mean $\boldsymbol{m}_{\text{MAP}}$.

3.2 Modeling Sensor Placement Decisions

Our interest, however, is not in only estimation but in optimal sensor placement locations. To this end we allow any of the spatial nodes $x_i,\ i = 0, 1, ..., N_x$ to be candidate sensor locations using the same discretization in x-direction as in §2.2. To allow the ability to select the position of the sensors, we associate with each x_i a non-negative binary weight $w_i \in \{0, 1\}$. Our intent is to denote by $w_i = 1$ the situation where a sensor is placed at location x_i and by $w_i = 0$ the situation where no sensor is placed at location x_i. Therefore, the problem of determining the optimal sensor locations becomes an large-scale mixed-integer integer nonlinear program. Our approach will be to perform relaxations of the sensor placement problem, by allowing w_i to have any value in in the domain $[0, 1]$.

We model the fact that a sensor has fixed spatial placement, at which we measure both flow and pressure at all times. Nevertheless, we allow grid points on the temporal direction to have the same weight. We thus create a weight diagonal matrix corresponding to each point in $\bar{\boldsymbol{u}}$:

$$W = \text{diag}(w_0, w_1, ..., w_{N_x}, w_0, w_1, ..., w_{N_x}, ..., w_0, w_1, ..., w_{N_x}) \in \mathbb{R}^{2m \times 2m}, \tag{3.25}$$

where $m = (N_x + 1)(N_t + 1)$ is the total number of discretized points in the domain $[0, T] \times [0, L]$.

Since in principle we allow any spatial degree of freedom to be measured, we initially assume that $\boldsymbol{f}(\boldsymbol{m})$ is the entire solution map $\mathcal{S}$, and we use the weights to winnow it down. Since we will end up solving an integer programming problem, we aim to produce a version of the optimal sensor placement that has a convex objective. Therefore, inspired by the workflow from [1], we approximate $\boldsymbol{f}(\boldsymbol{m})$ when used in (3.21) by its linearization around the prior mean $\boldsymbol{m}_{\text{prior}} = \bar{\boldsymbol{u}}_{0,prior}$. To this end, we denote the Jacobian of $\boldsymbol{f}$ at $\bar{\boldsymbol{u}}_{0,prior}$ by F. We then have that

$$\boldsymbol{f}(\boldsymbol{m}) \approx \boldsymbol{f}(\boldsymbol{m}_{\text{prior}}) + F(\boldsymbol{m} - \boldsymbol{m}_{\text{prior}}). \tag{3.26}$$

The last ingredient is to induce a weighted least squares setup to the estimation problem, to allow for a consistent statistical framework when allowing points to come in and out of the measurement set and thus decide on the optimal measurement set and sensor placement. This strategy is equivalent to scaling the variance of the measurement at a certain point x_i by $1/w_i$. For writing down the likelihood we need the inverse of the noise variance, which will now be $W^{0.5}\boldsymbol{\Gamma}_{\text{noise}}^{-1}W^{0.5}$. In this form, we assume that $\boldsymbol{\Gamma}_{\text{noise}}$ is the matrix of the noise as if sensors are at *every* grid point and are evaluated based on the covariance kernel (3.22). In this case (3.21) is proportional to the weighted least squares likelihood. That is, the w-weighted likelihood, conditional on the initial conditions $\bar{\boldsymbol{u}}_0$ and weights w, is [1]

$$\pi_{like}(d|\bar{\boldsymbol{u}}_0, w) \propto \exp\left\{-\frac{1}{2}(F\bar{\boldsymbol{u}}_0 - d)^T W^{\frac{1}{2}}\Gamma_{noise}^{-1}W^{\frac{1}{2}}(F\bar{\boldsymbol{u}}_0 - d)\right\}, \tag{3.27}$$

where d is a potential measurement at all times and space points. Accounting now for the prior distribution on $\bar{\boldsymbol{u}}_0$ around its mean $\bar{\boldsymbol{u}}_{0,prior}$, we get the posterior likelihood:

$$\pi_{post}(d|w) \propto \exp\left\{-\frac{1}{2}(F\bar{\boldsymbol{u}}_0 - d)^T W^{\frac{1}{2}}\Gamma_{noise}^{-1}W^{\frac{1}{2}}(F\bar{\boldsymbol{u}}_0 - d) - (\bar{\boldsymbol{u}}_0 - \bar{\boldsymbol{u}}_{0,prior})\Gamma_{prior}^{-1}(\bar{\boldsymbol{u}}_0 - \bar{\boldsymbol{u}}_{0,prior})\right\}. \tag{3.28}$$

We note that a maximum likelihood approach would have a similar expression except that it would miss the prior term. In that case the problem would become equivalent to one of least squares. Under the assumptions above, the distribution of the best estimate $\bar{\boldsymbol{u}}_0$ is normal with covariance matrix:

$$\Gamma_{post}(w) = \left(F^T W^{\frac{1}{2}}\Gamma_{noise}^{-1}W^{\frac{1}{2}}F + \Gamma_{prior}^{-1}\right)^{-1}. \tag{3.29}$$

3.3 Optimal Sensor Placement Formulation

The optimal sensor placement problem is cast as a design of experiments formulation. The aim is to minimize a measure of the posterior covariance matrix under the constraint of a fixed number of sensors. In this work we focus on minimizing the trace (corresponding to an A-optimality criterion) and minimizing the variance of the estimated initial flow. We capture these formulations using the following general form:

$$\min \quad \Psi(\Gamma_{post}(w)) \tag{3.30a}$$

$$\text{subject to} \quad w_i \in \{0,1\}, i = 0,1,...,N_x \text{ and } \sum_{i=1}^{N} w_i = n_0. \tag{3.30b}$$

Here n_0 is the total number of sensors to be placed. To minimize the posterior covariance of total flow, we use

$$\Psi\left(\Gamma_{post}(w)\right) \equiv a^T \Gamma_{post}(w)\, a, \quad a = (\underbrace{0,\ldots,0}_{N_x-1},\underbrace{1,\ldots,1}_{N_x+1}) \in \mathbb{R}^{2N_x}. \tag{3.31}$$

Here, the vector a has an entry of L/N_x corresponding to any initial flow variable, and 0 otherwise. In other words, the vector a is used to extract the flow variance from the posterior covariance matrix. The trace minimization problem considered in [1] uses

$$\Psi(\Gamma_{post}(w)) \equiv \text{Trace}(\Gamma_{post}(w)). \tag{3.32}$$

We can interpret the minimization of the trace of the posterior covariance as a compromise in aiming to reduce the variances of all possible linear functions of the initial state.

3.3.1 Sparsity-Inducing Approach

Because of the integrality of the sensor placement problem and the complex nonlinear structure of the measures used, direct use of off-the-shelf solvers does not result in scalable solutions. For instance, initial investigation using linearization of the mapping and use of mixed-integer linear solvers proved disappointing once we exceeded a mesh $N_x = 10$ (resulting in excessive compute times). To enable scalable solutions, we propose to use a sparse (compressed sensing) optimization approach [1] and a sum-up rounding approach [17]. In the compressed sensing approach, we introduce a sparsity-inducing penalty term while relaxing the binary constraints. This results in

$$\min \quad \Psi\left(\Gamma_{post}(w)\right) + \gamma \cdot \Phi(w) \tag{3.33a}$$

$$\text{subject to} \quad 0 \le w_i \le 1, i = 0,1,...,N_x \text{ and } \sum_{i=1}^{N} w_i = n_0. \tag{3.33b}$$

Here $\gamma \ge 0$ is a penalty parameter, and $\Phi(\cdot)$ is a penalty function. The ideal penalty function is the so-called 0-norm, which counts the nonzero entries. For γ sufficiently large, such a norm would indeed induce an integer solution. On the other hand, this formulation makes the problem difficult,

in effect NP-hard (in N_x). If the number of nonzero entries is small, however, an integer solution can be obtained with high probability by using the 1-norm (which is a continuous and convex metric). This is the basis for the recent advances in the area of compressed sensing [7]. If we insist on the constraint of the sum of weights being prescribed, however, then using $\Phi(w) = \|w\|_1$ has no effect on our problem. We have also tried to use $\Phi(w) = \|w\|_1$ without the total sum constraints $\sum_{i=1}^{N} w_i = n_0$, and chose the penalty parameter γ so that the solution of the problem satisfies $\sum_{i=1}^{N} w_i = n_0$. In the parameter ranges tried, this compressed sensing setup did not produce a sparse solution. An alternative is to use a penalty $\Phi(w)$ that is closer to the 0-norm, although this comes at the cost of abandoning convexity. Such an approach is also used in [1]. In this work we use $\Phi(w) = \|w\|_{1/2}$. We highlight that the cost function $\Psi\left(\Gamma_{post}(w)\right)$ may be nonconvex and the penalty term $\Phi(w) = \|w\|_{1/2}$ would add to nonconvexity.

3.3.2 Sum-up Rounding Approach

The other approach considered is the sum-up rounding (SUR) strategy of Sager. This approach starts with the convex relaxation of the optimization problem (3.30) (and formally represents the problem (3.33) for $\gamma = 0$):

$$\min \quad \Psi\left(\Gamma_{post}(w)\right) \tag{3.34a}$$

$$\text{subject to} \quad 0 \leq w_i \leq 1, i = 0, 1, ..., N_x \text{ and } \sum_{i=1}^{N} w_i = n_0. \tag{3.34b}$$

An important property is that this problem produces a lower bound for the objective function of (3.30). The key is now to produce an upper bound by finding an integer vector w that satisfies the constraint $\sum w_i = n_0$ and that has an objective only slightly increased from the optimal one of (3.34). In the sum-up rounding approach, an upper bound is produced as follows. If we denote the relaxed solution of (3.34) by $\boldsymbol{w_{rel}} = \{w^0_{rel}, ..., w^{N_x}_{rel}\}$, then an integer-valued solution $\boldsymbol{w_{int}} = \{w^0_{int}, ..., w^{N_x}_{int}\}$ is obtained by:

$$w^j_{int} = \begin{cases} 1 & \text{if } \sum_{k=0}^{j} w^j_{rel} - \sum_{k=0}^{j-1} w^j_{int} \geq 0.5 \\ 0 & \text{otherwise} \end{cases} \tag{3.35}$$

for $j = 0, 1, ..., N_x$. It has been shown for optimal control problems with binary controls [18] and for the optimal selection of measurement times in time-dependent initial value problems [17], that this rounded solution will become arbitrarily close to the relaxed one when the underlying grid is chosen fine enough. The problem we apply this strategy to is different, because rounding occurs in space and in the presence of two-point boundary conditions. Hence, it is unclear wheter such desirable properties persist. We investigate this situation in the following section.

4 Numerical Results

We now present numerical results of approximations to the design of experiment problems with the methods described above. We consider the case of one pipeline for which the physics is discretized by the upwind scheme (2.15) and for which the design optimization problem (3.30) is set up.

One element that we need to specify for (3.30) is the choice of mean of the prior distribution, $\bar{\boldsymbol{u}}_{0,prior}$, which is used, for example, as the point at which the nonlinear mapping $\boldsymbol{f}$ from (3.17) is linearized (3.26).

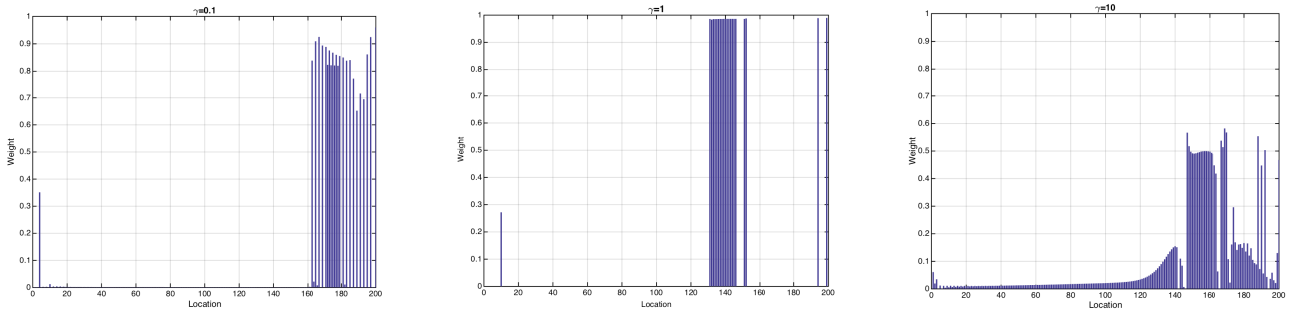
We obtain that point in the following way. After some time, the dynamics of pressure and flow in the pipeline become stationary, and we use that value as the prior mean. The diagonals of both prior and noise covariance matrices are chosen to be 1; the average mean prior flow is about 10, and pressure decreases from 10 to 1. We note that the absolute value of the covariance does not influence the design but changing the ratio of sensor noise and prior covariance does. The proper value depends on the situation being modeled, and for this initial investigation we decided to choose them to be comparable. The relationship between the sensor noise and mean observation, however, is a function of the sensor rating and does carry physical significance. This indicates a ratio of the noise to the mean observation of about 10% (assuming flow value ranges being similar to the mean). We note that this value is in the range of existing sensors for typical flow regimes [13]. We experimented with smaller relative noise variances; and when we applied the sum-up rounding method for that case, we observed similar gap reductions (reported below) so in that sense the performance of the method is similar. The designs, however, approach *uniform designs* so they do not appear as interesting. We thus carry out the computation with the parameters specified above. With this setup, we investigate the variance minimization problem and the trace minimization problem. All the calculations were performed in Matlab.

4.1 Case I: Total Flow Variance Minimization

The optimal design of experiments framework is first applied to the total initial flow variance minimization problem given by (3.30) with the objective function (3.31). In this case study we compare the sparsity-inducing optimization approach and the sum-up rounding strategy discussed in §3.3.

4.1.1 Sparsity-Inducing Approach

For the variance minimization problem we use $L = 80000$, $T = 60$ based on [24]. Also, using the quantities from [24], we obtain the following parameters. $c_1 = 9.6917$, $c_2 = 14137$, $c_3 = 0.0825$. To obtain a stable discretization, we must obey the CFL restrictions. We thus choose $N_x = N_t = 200$. We then solve the problem (3.33) using the `fmincon` optimization routine in Matlab. We choose as a starting point for the weights $w_i = 0.5$ for all i. We solve the optimization problem (3.33) using a constrained optimization solver. The results for white noise are displayed in Figure 1 for increased values of γ. We see that increased γ indeed promotes increased sparsity and a (mostly) integer solution. On the other hand, the changes in the solution for increased γ are large and irregular. We thus



(a) $\gamma = 0.1$ (b) $\gamma = 1$ (c) $\gamma = 10$

Figure 1: Optimal placement for sparsity-inducing (compressed sensing) approach (3.33). Flow variance minimization formulation.

suspect that the sparsity-inducing solution may not get close enough to the actual solution, but we do not have enough evidence one way or the other to make a final conclusion.

4.1.2 Sum-Up Rounding Approach

We use $L = 80000, T = 60, c_1 = 9.70, c_2 = 1.41 \times 10^4, c_3 = 8.25 \times 10^{-2}, f_0(x) = 10, p_0(x) = -9x/L + 10$, and $\tau_1 = 2, \tau_2 = 6$ in Figures 2, 3, and 4. The time constant governing the dynamic response of some flowmeters is about 1 second [23]; this governs our choice of τ_1. We chose $\tau_2 = 3.0 \cdot \tau_1$ both to capture enough of the dynamic response of the sensor and to obtain a reasonably sparse covariance matrix to help with computations. We present the gap between the objective function of the relaxation (3.34) and the sum-up rounding integer value (3.35). We present the gap for white noise and colored noise cases and the gap scaled by the mesh size. Scaling here was necessary to compare problems of different sizes and to validate that the gap was shrinking with increasing mesh resolution. While the solution does not change, for the objective function to represent total variance, the vector a in (3.31) must be scaled by the mesh size, which varies with $1/N_x$. The lower and upper bounds in Figures 2 and 3 are not scaled in order to aid visualization.

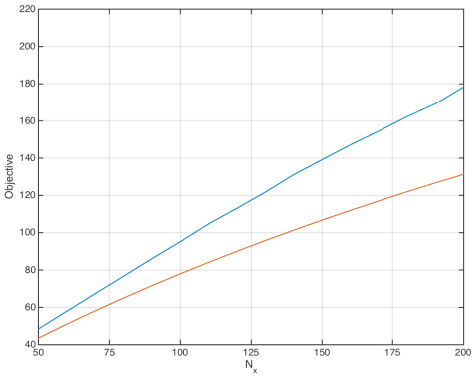


Figure 2: Sum-up rounding upper/lower bounds (white noise case).

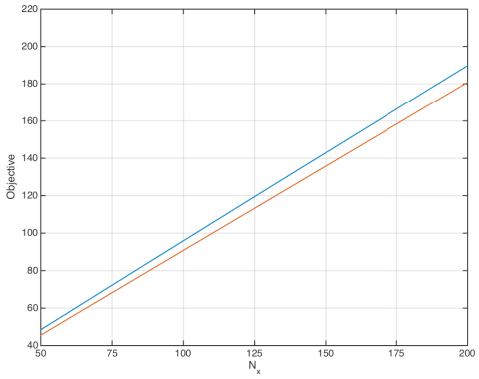


Figure 3: Sum-up rounding upper/lower bounds (colored noise case).

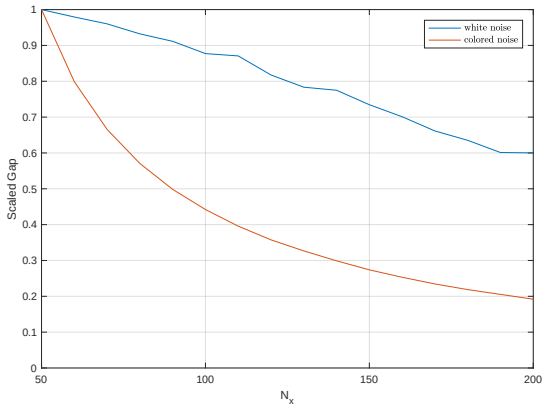


Figure 4: Sum-up rounding gap percentage

In Figure 5 (optimal sensor locations), the grid size used is 100×100 and $n_0 = 10$, which means that we have 100 candidate locations. These candidate locations are evenly distributed along the pipeline, but only 10 are selected to place sensors. In Figure 4, the y axis is the percentage of the gap in the objective function when N_x and N_t are determined by x relative to the gap when $x = 1$. In other words, we want to see how the gap decreases as the mesh gets more refined, since this seems to be the message in [18, 17]. The number of sensors is kept to a fixed fraction of the number of grid points, in this case $n_0 = 0.1 \times N_x$.

Given the small magnitude of the gap, we believe that the results for the optimal distribution of the sensors are reasonable. We note that the sensor distributions, while not complex, are certainly not trivial as they show that the spacing increases from inflow to outflow in the white noise case and that it reaches the maximum spacing two thirds away from the inflow in the colored noise case. We also observe that using colored noise against white noise results in a significantly smaller gap.

We also computed the total variance at the solution of the sparsity-inducing approach, and it was indeed worse. For example, for $N_x = 200$, it was worse by about 4.65%. Figure 1 looks much worse than that; part of the problem occurs because the total flow variance is not very sensitive to sensor placement if the time horizon is short, here being 60 s. This short time horizon was chosen because of computational time constraints. In any case, we conclude that the sparsity-inducing approach exhibits inferior performance compared with that of the sum-up rounding approach.

4.2 Case II: Trace Minimization

In Figure 6, the solution of §4.1.1 in the case of white noise is presented. The conclusions are similar to the ones obtained by minimizing the variance of the total flow. The placement appears to be unstable as γ is varied, and subsequent comparisons with the solution of the sum-up rounding approach convinced us that the solution is far from optimum. For $N_x = 100$, the solution produced is worse by 7.5% in terms of the value of the objective (the trace of the posterior covariance). We observed a similar behavior for the colored noise approach in the few circumstances studied.

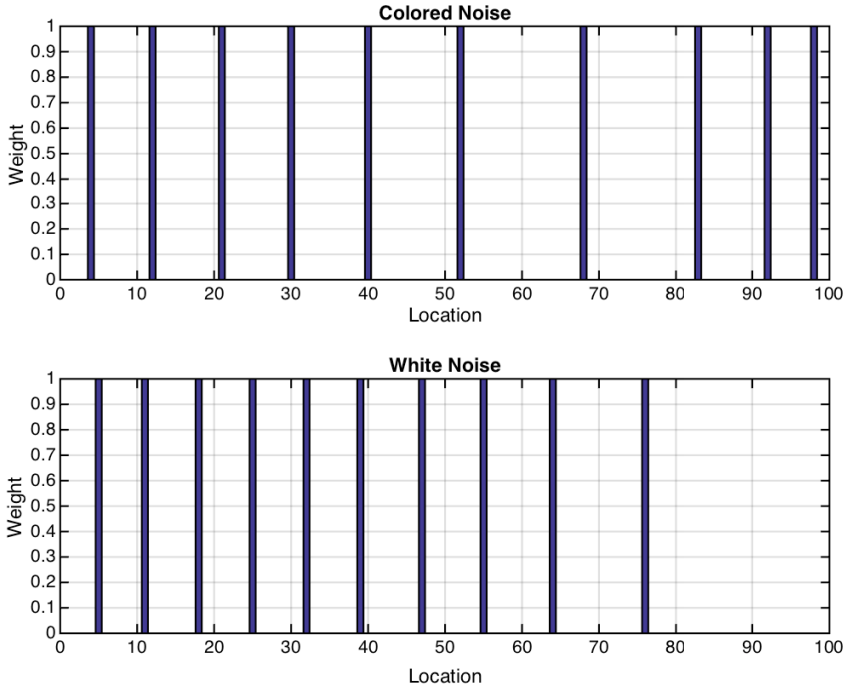
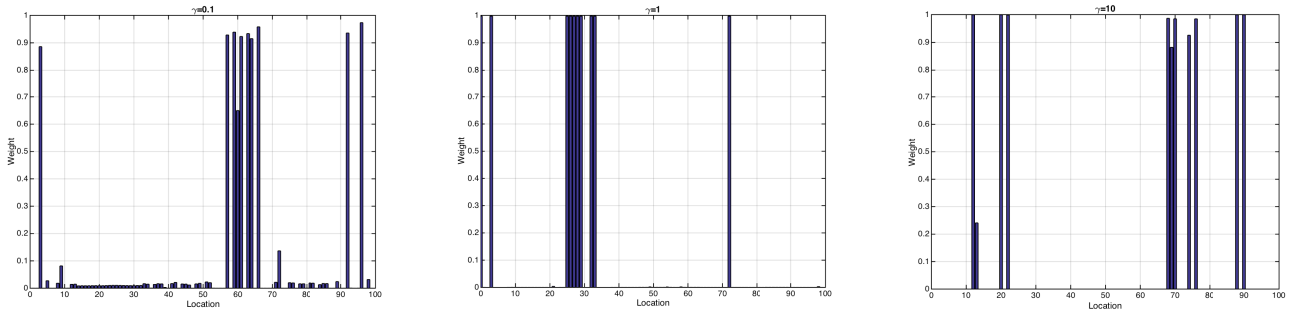


Figure 5: Optimal sensor locations using the sum-up rounding approach in the total flow minimization example.

We thus focus on applying the sum-up rounding approach to the A-optimal experimental design problem (3.30) that minimizes the trace of the covariance matrix (3.32). That is, the relaxed problem (3.34) is solved followed by the sum-up rounding strategy (3.35) to produce the integer solution. Specifically, suppose w_{rel} is the solution to the relaxation (3.33) and w_{int} is the integer solution obtained from w_{rel} via the rounding-up strategy (3.35). Recall that the posterior covariance matrix is given by (3.29).

We use the notation $C_1 = \Gamma_{post}(w_{rel})$ and $C_2 = \Gamma_{post}(w_{int})$. How well we can certify that the sum-up rounding strategy works depends on how close C1 and C2 are to each other. In keeping with the intuition behind the sum-up rounding method [18], it is expected that the gap decreases as the mesh is refined, provided that the number of sensors are kept at a constant proportion of the number of nodes. One challenge faced here, however, is comparing the optimality gap for different problem instances. In the case of the variance of total flow, described in §4.1, this is a discretized linear form of the solution; the weights from the discretization indicate what the scaling should be, so that one can reason whether the gap is large or small from a practical perspective. The trace of the posterior variance does not naturally represent an observable quantity that can be expressed as an integral objective and in this sense this problem does not immediately fit the setting used in [18]. Because of this, the sum-up rounding strategy is used only as a heuristic. An example of the difficulty is the following. Each diagonal entry in C_i, for a fixed number of sensors, would have to converge to a fixed value (the variance of the flow at the given point), but its trace would go to infinity. To this end, the gap $C_1 - C_2$ is mapped to metrics that can be more easily interpreted. One alternative is to compute Trace$(C_1 - C_2)$ scaled by the number of points in the x direction, which is equivalent to multiplying

with ΔX. Another option is to compute the largest eigenvalue of $C_1 - C_2$ also scaled by ΔX, we call this quantity the *largest eigenvalue distance*. Another option is to compute the difference between the variances for the total flow, that is, $a^T(C_2 - C_1)a\left(\frac{L}{N_x}\right)^2$, where a is the vector in the objective function of total flow. We call the absolute value of this quantity the *total flow variance distance*. In this last case, we can compare this gap with the gap from the total flow variance minimization problem §4.1.



(a) $\gamma = 0.1$ (b) $\gamma = 1$ (c) $\gamma = 10$

Figure 6: Optimal placement for sparsity-inducing (compressed sensing) approach (3.32) for the trace minimization formulation with white noise.

We decided not to present the results for the *white noise* case because this assumption is unrealistic and the resulting figures present behaviors that are similar to those of the *colored noise* case. Moreover, as in the total flow minimization case, they exhibit a slower decrease in the gap compared with the colored noise case.

For *colored noise*, with the measurement error covariance kernel described in (3.22), our results are presented in Figures 7–12.

- The resulting upper and lower bounds are presented in Figure 7 (scaled by $\frac{1}{n_p}$ in Figure 8) and the gap between them is presented in Figure 9. The x axis is $N_x = N_t = n_p$ (the number of parameters), the y axis is $\text{Trace}(\Gamma_{post}(w_{int}))$.
- The best sensor configuration (from the sum-up rounding strategy for $N_x = 100$) is presented in Figure 10.
- The largest eigenvalue distance (the eigenvalue with the largest absolute magnitude of $(C_2 - C_1)\Delta X$) is displayed in Figure 11. The total flow variance distance is presented in Figure 12 (this is equivalent to computing $a^T(C_2 - C_1)a\left(\frac{L}{N_x}\right)^2$ with the notation from (3.31)). On the x axis we display $N_x = N_t$.

From the numerical experiments, we conclude the following. As N_x increases, the gap shrinks. The most convincing evidence we find is the largest eigenvalue discrepancy plots in Figure 11, particularly when corroborated with the scaled upper and lower bounds calculations from Figure 8. Note that despite the fact that $\text{Trace}(C_2 - C_1)\Delta X$ is almost constant, in the spectral norm $(C_2 - C_1)\Delta X$ is decreasing significantly, and the rate is faster than $\frac{1}{\sqrt{N_x}}$ (Figure 11). The scaled gap, in particular, does not appear to converge (Figure 9). This also shows that, from the comparison with the largest eigenvalue distance behavior, scaling the trace by its size is not the right approach. The right one will

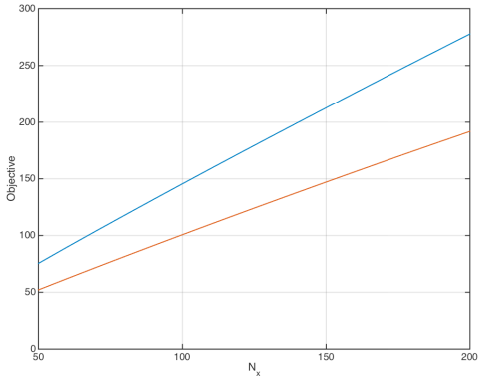


Figure 7: Upper/lower bounds for trace minimization case.

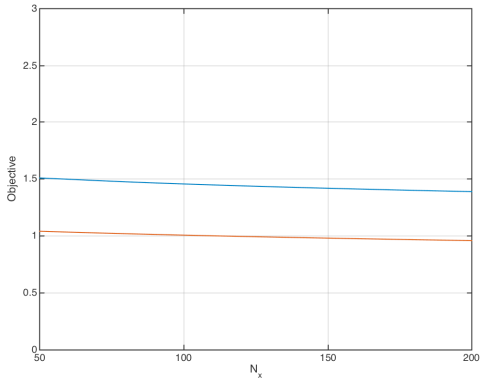


Figure 8: Scaled upper/lower bounds for trace minimization case.

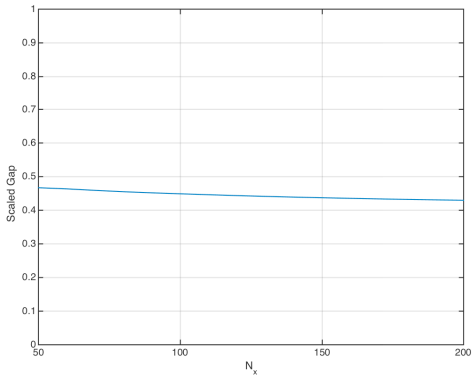


Figure 9: Scaled gap for colored noise, trace minimization case.

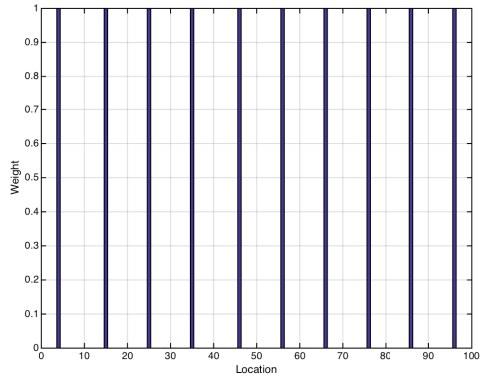


Figure 10: Optimal sensor placement for colored noise, trace minimization case.

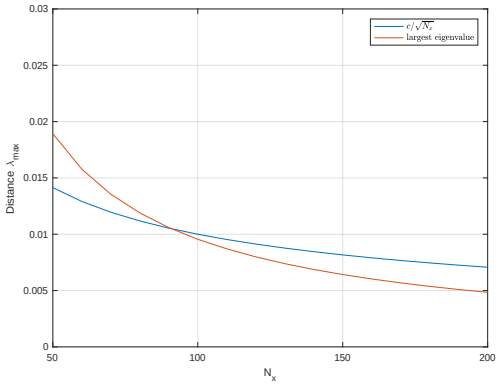


Figure 11: Largest eigenvalue distance between relaxed and sum-up rounding covariances for colored noise, trace minimization case.

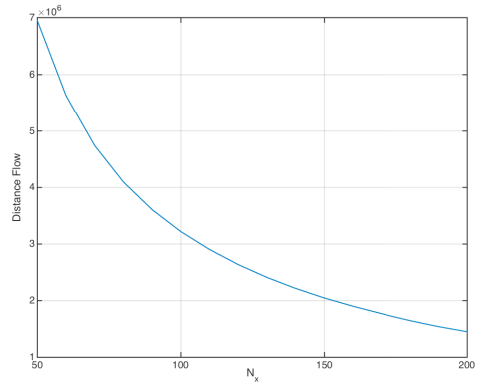


Figure 12: Total flow variance distance between relaxed and sum-up rounding covariance for colored noise, trace minimization case.

be the subject of future research and is an intriguing problem in itself. We also display the total flow variance distance in Figure 12. In this case, the proper scaling is clear, since it represents differences

of statistics of physical random variables. We observe a decrease that is comparable to that in the total flow minimization optimization case. Specifically, we have a decrease by a factor of 5, between the case $N_x = 50$ and $N_x = 200$, whereas in this trace minimization case we have a decrease by a factor of 4.5 between the same mesh sizes in the total flow variance distance.

The plots for the optimal solution, shown in Figure 10, indicate that a uniform distribution is a reasonable approximation. While the patterns show some spatial variability, it is small compared with the gap to indicate that the exact solution would not be uniform or close to it. In that sense this approach does not bring new insights, with the exception that the approach indicates that the optimal placement will not include samples at the endpoints (again, an otherwise understandable conclusion given such results from approximation theory). On the other hand, the endpoints are places where in practice the industry will most likely have sensors. Thus, the second iteration of this may be to assume that sensors exist at the endpoints and to solve the new problem. Our guess is that the result will be close to uniform distribution again. Our interpretation of such effects comes from the fact that minimizing the trace of the covariance is a bit like minimizing for the variance of all possible linear forms of the initial state (in effect, one can prove this is precisely the average of variances over random uniform choices of a over a sphere). With this interpretation it is perhaps not surprising that a close-to-uniform distribution ensues. As a check suggested by a referee, we tried each best solution we found for the total variance minimization and the trace minimization setups in the other problem, and we found that they had larger objective values, but not by much (less than 1%). Therefore, in the parameter setup we tried, the minima appear to be shallow, and uniform designs appear to be acceptable. We suspect that an issue may be the limited time horizon we can consider because of the intensive nature of the computation (it takes about a day to solve the problem in Matlab for 200 mesh points and a 60-seconds horizon on a laptop). More extensive simulations are needed to elucidate such initial questions.

5 Final Discussion and Conclusions

We have presented an approach for the scalable computation of the optimal design of experiments and its application to sensor placement in gas pipelines. The objective of such an endeavor is to reduce the uncertainty in the quantities being estimated. We are particularly interested in the consequences of integration of natural gas and electricity, which is an issue of growing interest. As one of us demonstrated in a recent paper [24], an increase in renewable energy reliance for electricity would bring increased changes in usage in ways that do not conform with existing usage patterns, so that the available gas may need to be better estimated.

To this end, we have proposed a Bayesian framework in which we compute the posterior covariance of initial condition given the sensor positions, and then we design the sensor placement to reduce a measure of this covariance. A key computational and modeling effort was to compute the parameter to data mapping, something we have done recognizing the hyperbolic nature of the problem and positing a stable discretization of that operator. We have considered both white noise distribution of the measurements (a conservative assumption) and colored noise (a more realistic

one). We have also considered two criteria for optimal sensor placement: total flow variance and A-optimal design. The resulting problem is a mixed-integer nonlinear program. To prepare for finer mesh and multidimensional PDE-based problems, we looked at solvers that have the potential to scale well for such large-scale problems, which typically classical MINLP solvers cannot do. To this end, we experimented with a recently proposed sparsity-inducing (compressed sensing) approach (which has shown satisfactory results results in geoscience applications [1]) as well as with a sum-up rounding approach [18, 17].

5.1 Discussion

We concluded that in our setup the sparsity-inducing approach did not produce an integer solution of sufficient quality. Because we do not know the exact solution, we came to this conclusion by examining the stability of the approach with respect to the shrinking parameter and by comparing it with the results of the sum-up rounding strategy. We observed that the latter produced good results particularly in the increasing discretization case, as expected because of the interpretation of the approach as being provably optimal in the limit of increasingly accurate discretizations of a continuous limiting case [18]. For the total flow variance minimization case, the gap in the relaxation and the sum-up rounding integer solution decreases roughly as $\frac{1}{\sqrt{N_x}}$, where N_x is the number of points in the x direction. This indicates indeed an increasingly accurate solution.

For the A-optimal design (the trace minimization problem) the comparison is more complicated, because classical design of experiments theory has a framework for increasing the number of observations but not for increasing the dimension of the problem in terms of comparing the objectives of solutions for different problem sizes. We experimented with different comparison metrics, the most satisfying of which we found to be the largest absolute value of the eigenvalue of the difference between the covariance of the relaxed solution and that of the integer-valued problem. We found that in this metric the sum-up rounding strategy also results in decreasing the optimality gaps with problem size in the same $\frac{1}{\sqrt{N_x}}$ fashion.

Comparing the white noise with the colored case, we observed that the solution of the white noise case is harder for both total flow variance minimization and A-optimal design, in the sense of the metrics of convergence being larger and slower to converge, although not significantly so. Since the colored-in-time noise case is more realistic for large data (even if the white noise case is more common in the literature), this is a fortunate occurrence in our opinion. A good outcome for white noise situations is also not conceptually covered by the 1D ansatz in [18], since in this case the embedding problem is closer to a two-dimensional one. But the optimality gap shrinks as the mesh is refined in the white noise case as well.

In the total flow variance case the optimal sensor placement solution departs significantly from the uniform distribution, whereas in the A-optimal design the optimal solution appears close to uniform. While our approach presents the first computational evidence of this fact for our target problem class, it is certainly disappointing, although given the worst-case flavor of the A-optimal design perhaps not entirely unexpected. Of course when experimenting with different regimes this conclusion may change, but this is how the evidence sits at the moment.

While many more experiments need to be run, the behavior of the total flow optimal sensor placement suggests that the optimal solution while focusing on target functionals of the unknown flow may result in nontrivial geometrical patterns of sensor placement. Of course this often may not be desirable because the objective functional may change with usage and redeployment of sensors may be difficult. Therefore, a conservative approach would be the A-optimal design one. But when desired, it appears to be an interesting direction of investigation in terms of expected outcome.

5.2 Future Work

To summarize our conclusions from this initial investigation, we found that sum-up rounding produces good-quality integer solutions in the limit of fine meshes, that the sparsity-inducing compressed sensing approach does not work well for this problem, that colored noise problems appear to be easier than white noise ones, and that a narrow focus on the objective of the problem may result in nontrivial (distinct from uniform) sensor distribution patterns.

Future research needs to extend this work in several directions. The current solution setting is limiting: the larger optimization problems require solution times of one day on a laptop. More efficient implementations will be sought, which will also allow us to compare performance with other strategies such as those implemented in `casiopeia` (`https://github.com/adbuerger/casiopeia`). Some sensors may exist in the system by construction (such as the ends of the pipeline) or limited sensor measurements (e.g., only flow) may be available. Since the methodology presented is general, these other settings can be explored. Another intriguing question is the proper scaling of the A-optimal design between problems of different size, in order to compare the behavior of the relaxation for increasing mesh size, an important requirement of embedding the problem in a sum-up rounding strategy. An important modeling enhancement that would increase the value of the approach for real systems are the usage of a rolling horizon approach as opposed to a one-shot estimation approach.

Acknowledgment

We are grateful to Noemi Petra for suggestions at the beginning of this work. Jing Yu and Mihai Anitescu were supported by the U.S. Department of Energy, Office of Science, under contract DE-AC02-06CH11357. Victor M. Zavala acknowledges funding from the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison and from the industrial members of the TWCCC consortium.

References

[1] A. Alexanderian, N. Petra, G. Stadler, and O. Ghattas. A-optimal design of experiments for infinite-dimensional Bayesian linear inverse problems with regularized l_0-sparsification. *SIAM Journal on Scientific Computing*, 36(5):A2122–A2148, 2014.

[2] J. Berry, W. E. Hart, C. A. Phillips, J. G. Uber, and J.-P. Watson. Sensor placement in municipal water networks with temporal integer programming models. *Journal of Water Resources Planning and Management*, 132(4):218–224, 2006.

[3] J. W. Berry, L. Fleischer, W. E. Hart, C. A. Phillips, and J.-P. Watson. Sensor placement in municipal water networks. *Journal of Water Resources Planning and Management*, 131(3):237–243, 2005.

[4] L. Billmann and R. Isermann. Leak detection methods for pipelines. *Automatica*, 23(3):381–385, 1987.

[5] R. G. Carter, T. F. Dupont, and H. H. Rachford Jr. Pack management and transient optimization of natural gas transmission lines. In *IGRC Conference, Vancouver*, 2004.

[6] D. J. Chmielewski, T. Palmer, and V. Manousiouthakis. On the theory of optimal sensor placement. *AIChE Journal*, 48(5):1001–1012, 2002.

[7] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

[8] C. M. Harris, K. L. Hoffman, and L.-A. Yarrow. Using integer programming techniques for the solution of an experimental design problem. *Annals of Operations Research*, 58(3):243–260, 1995.

[9] J. Kaipio and E. Somersalo. *Statistical and Computational Inverse Problems*, volume 160 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2005.

[10] A. Krause, J. Leskovec, C. Guestrin, J. VanBriesen, and C. Faloutsos. Efficient sensor placement optimization for securing large water distribution networks. *Journal of Water Resources Planning and Management*, 134(6):516–526, 2008.

[11] A. Krause, A. Singh, and C. Guestrin. Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies. *The Journal of Machine Learning Research*, 9:235–284, 2008.

[12] R. M. Mattheij, S. W. Rienstra, and J. H. ten Thije Boonkkamp. *Partial Differential Equations: Modeling, Analysis, Computation*. SIAM, 2005.

[13] M. Menezes. Calculating & optimizing repeatability of natural gas flow measurements. Technical Note 00840, Emerson Process Management, 2012.

[14] E. Musulin, C. Benqlilou, M. J. Bagajewicz, and L. Puigjaner. Instrumentation design based on optimal Kalman filtering. *Journal of Process Control*, 15(6):629–638, 2005.

[15] H. H. Rachford Jr, R. G. C. Advantica, and T. F. D. Advantica. Using optimization in transient gas transmission. In *PSIG Annual Meeting*. Pipeline Simulation Interest Group, 2009.

[16] H. P. Reddy, S. Narasimhan, and S. M. Bhallamudi. Simulation and state estimation of transient flow in gas pipeline networks using a transfer function model. *Industrial & Engineering Chemistry Research*, 45(11):3853–3863, 2006.

[17] S. Sager. Sampling decisions in optimum experimental design in the light of Pontryagin's maximum principle. *SIAM Journal on Control and Optimization*, 51(4):3181–3207, 2013.

[18] S. Sager, H. G. Bock, and M. Diehl. The integer approximation error in mixed-integer optimal control. *Mathematical Programming*, 133(1-2):1–23, 2012.

[19] A. K. Singh and J. Hahn. Sensor location for stable nonlinear dynamic systems: Multiple sensor case. *Industrial & Engineering Chemistry Research*, 45(10):3615–3623, 2006.

[20] D. Telen, F. Logist, R. Quirynen, B. Houska, M. Diehl, and J. Impe. Optimal experiment design for nonlinear dynamic (bio) chemical systems using sequential semidefinite programming. *AIChE Journal*, 60(5):1728–1739, 2014.

[21] J. Van Deen and S. Reintsema. Modelling of high-pressure gas transmission lines. *Applied Mathematical Modelling*, 7(4):268–273, 1983.

[22] J.-P. Watson, H. J. Greenberg, and W. E. Hart. A multiple-objective analysis of sensor placement optimization in water networks. In *Proceedings of the World Water and Environment Resources Congress. American Society of Civil Engineers*, 2004.

[23] D. Wiklund and M. Peluso. Quantifying and specifying the dynamic response of flowmeters. *TECHNICAL PAPERS-ISA*, 422:463–476, 2002.

[24] V. M. Zavala. Stochastic optimal control model for natural gas networks. *Computers & Chemical Engineering*, 64:103–113, 2014.